

6.4 Klasyfikator SVM

Zaimplementuj uczenie liniowej maszyny wektorów nośnych w oparciu o metody optymalizacji numerycznej dostępne w scipy oraz funkcję nagrody:

$$L(\vec{w}, b) = \sum_i \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j \vec{x}_i^T \vec{x}_j \quad (5)$$

- $\lambda_n \in \mathbb{R}_{\geq 0}$ waga n -tego wektora nośnego
- $\vec{x}_n \in \mathbb{R}^D$ n -ta próbka ze zbioru uczącego
- $y_n \in \mathbb{R}$ n -ta etykieta ze zbioru uczącego

Wytrenuj klasyfikator na **syntetycznych jednomodowych zbiorach** danych i porównaj jego działanie z `sklearn.svm.SVC` z `kernel='linear'`.

Zaimplementuj uczenie maszyny wektorów nośnych wykorzystującej radialną i wielomianową funkcję jądra(6) i porównaj skuteczność i kształt granicy decyzyjnej z implementacją z `sklearn` z `kernel='rbf'` i `kernel='poly'` na syntetycznych zbiorach dwuwymiarowych danych z jednym klastrem na klasę.

$$L = \sum_i \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j K(\vec{x}_i, \vec{x}_j) \quad (6)$$

Wykorzystując zbiór Stellar Classification Dataset - SDSS17 (© SDSS) i algorytm selekcji modelu `sklearn.model_selection.GridSearchCV` dobierz optymalny zestaw hiperparametrów (`C` i `kernel`) i naucz model SVM (z `sklearn`) do klasyfikacji obiektów jako gwiazdy lub galaktyki. Przed doбором hiperparametrów uzupełnij brakujące wartości (jeżeli istnieją) i wyeliminuj nieistotne kolumny ('cam_col', 'MJD' oraz te kończące się na '_ID'), narysuj macierz korelacji pomiędzy wartościami i dokonaj normalizacji wartości. Do jakich wartości macierzy pomyłek należy dążyć jeżeli chcemy wykorzystać klasyfikator do automatycznej selekcji obiektów do obserwacji teleskopem podczas poszukiwania nowych galaktyk?

6.4.1 Kryteria oceniania

Implementacja liniowego SVMa $\rightarrow 3$

Implementacja SVMa wykorzystującego funkcję jądra $\rightarrow 4$

Analiza zbioru SDSS17 w oparciu o SVM $\rightarrow 5$